Data driven safe vehicle routing analytics: a differential evolution algorithm to reduce CO2 emissions and hazardous risks

Article (Accepted Version)

Data driven safe vehicle routing analytics: A differential evolution algorithm to reduce CO_2 emissions and hazardous risks

Boon Ean Teoh
Monash University Malaysia,
Bandar Sunway, Selangor, Malaysia
Email: teoh.boon.ean@monash.edu

S.G. Ponnambalam
Advanced Engineering Platform and School of Engineering,
Monash University Malaysia,
Bandar Sunway, Selangor, Malaysia
Email: sgponnambalam@monash.edu

Nachiappan Subramanian*
School of Business Management and Economics,
University of Sussex,
Jubilee Building, 204, Falmer, Brighton BN1 9SL.
N.Subramanian@sussex.ac.uk
Phone: +44 1273 872982

*** Corresponding Author**

Submitted to the Annals of Operations Research special issue titled "Big Data Analytics (BDA) in Operations and Supply Chain Management"
.

Data driven safe vehicle routing analytics: A differential evolution algorithm to reduce CO_2 emissions and hazardous risks

Abstract:

Contemporary vehicle routing requires ubiquitous computing and massive data in order to deal with the three aspects of transportation such as operations, planning and safety. Out of the three aspects, safety is the most vital and this study refers safety as the reduction of CO_2 emissions and hazardous risks. Hence, this paper presents a data driven multi-objective differential evolution (MODE) algorithm to solve the safe capacitated vehicle routing problems (CVRP) by minimizing the greenhouse gas emissions and hazardous risk. The proposed data driven MODE is tested using benchmark instances associated with real time data which have predefined load for each of the vehicle travelling on a specific route and the total capacity summed up from the customers cannot exceed the stated load. Pareto fronts are generated as the solution to this multi-objective problem. Computational results proved the viability of the data driven MODE algorithm to solve the multi-objective safe CVRP with a certain trade-off to achieve an efficient solution. Overall the study suggests 5% increment in cost function is essential to reduce the risk factors. The major contributions of this paper are to develop a multi-objective model for a safe vehicle routing and propose a multi-objective differential evolution (MODE) algorithm that can handle structured and unstructured data to solve the safe capacitated vehicle routing problem.

Keywords: Safe capacitated vehicle routing; Greenhouse gas emission; Hazardous risk; Multi-objective Differential evolution

1 Introduction

Primary role of logistics in the digital world is to use real time data to generate a vehicle routing which reduces negative consequences such as congestion, safety and environment. If logistics is not managed well, it will cause congestion and enormous greenhouse gas (GHG) emissions (Savelsbergh and Van Woensel 2016; Hazen et al. 2016). Recently, logistics and distribution systems have been identified as one of the most expensive aspects for an organization to deal with the ecological impact which is one of the biggest challenges (Savelsbergh and Van Woensel 2016; Sinuany-Stern and Sherman 2014). Freight GHG emissions increased by 46% from 1990 to 2003 according to United States Environmental Protection Agency (EPA 2006) and therefore this is an emergent need to reduce the amount of emission from this sector.

Transportation is a major cause for environment degradation in the modern world (Rodrigue et al. 2001). Vehicles consume energy in the form of oil and emit pollutants such as carbon dioxide (CO_2) which contributes to the greenhouse gasses (GHG) emissions. In the era of e-commerce, increased mobility due to personalised demands for customers residing in the urban area results in more congestion (Rodrigue 2013). In particular, transportation industry is subjected to cause severe environmental problems such as climate change and air pollution (Brandenburg and Rebs 2015). Other than transportation of regular materials, there is a need for cautious planning and scheduling to transport hazardous materials gasses, explosives and flammables from source to destination using big data. For instance, the oil and gas industry transports the oil extracted from the plants to their manufacturing sites and then distribute it to storage tanks across the country. During transportation, the possibility of an accident may pose a health threat to the nearby population and property.

Therefore, the task of scheduling and planning for safe logistics management in the digital era is vital and this is achievable with the proper use of enormous amount of information that is available either structured or unstructured. It is quite obvious that big data can support supply chain managers to develop an optimal route that will not only reduce cost but also reduce CO_2 emissions. Optimization of the routes will ensure that the vehicles are travelling on lower risk track which ultimately reduce public risk from the shipments of these materials. Hence safe route planning will involve deciding on the route to be taken and the timing which is suitable in order to reduce the time spent on road.

Safe routing and planning becomes more complex in the case of hazardous materials to reduce ecological impact. Major hazardous material accidents have been reported and although they are not catastrophic, but the causality of the accident are severe that affected the neighbourhood. A recent hazardous material accident in Mexico which involves the ammonia truck resulted into 39 causalities (Verter 2011). Hence, various agencies came out with several regulations to reduce such incidents in which one among them is not to use roads which are highly populated for vehicles that carry dangerous materials. The transportation risk due to an accident will affect the residents living or working around the area. This gives rise to the emergency response plan that is put into use to reduce the effect of a hazardous material accident.

The safe plan includes establishing a team with better coordination and also includes those who are specialized in handling the respective material. However, prevention and minimization of the accident risk is at higher priority. Given the availability of data, how to effectively use these various forms of real time structured and unstructured data to develop a safe vehicle routing analytics that minimises CO_2 emissions and risk remains an open research question.

The paper is organised as follows where section 2 reviews previous studies on safe transportation and the role of big data in logistics. Safe vehicle routing optimisation modelling that considers CO_2 emissions and risk are explained in section 3. Data driven multi objective algorithm is explained in section 4. Robustness of data driven MODE is evaluated in terms of parameter fine tuning and solution quality and finally the conclusion summarises the work with further research directions.

2 Literature Review

2.1 Big data and logistics management

Digital universe throws unlimited access to different forms of data such as photographs, surveillance video feeds; and data produced through sensors challenges logistics industry on how to make use of these real time data to develop intelligent and safer transportation (OECD 2015). Intelligent transport refers to visualisation and analysis of real time usage of transport network and safer transport refers to processing of real time data with respect to vehicle operation and to protect the surrounding environment to avoid or minimise potential dangerous conflicts (OECD 2015). There is a huge potential for researchers to come out with safe models and algorithms using the availability of data to support the policy makers to develop new regulations to reduce congestion and increase safety. For example, real time data collected through several gantry cranes erected on freeways to monitor the vehicle usages with eTags in Taiwan made automatic toll-collection that substantially reduced CO_2 emissions, travel time and congestion. Hence, big data can very well support logistics researchers to develop safer vehicle routing models and analytical methods.

2.2 Vehicle routing models and methods to reduce CO_2 emission

CO_2 emissions from transportation accounted for one of the highest percentage as compared to other sectors as shown in the US environmental report (EPA 2014). Davis et al. (2005) proposed a vehicle emissions model which estimates vehicle emissions in any area with given inputs. With information of the vehicle fleet, the model proposed can be used to predict total emissions. Changes in emission can also be detected if there is a change in fleet, fuel and congestion. This model is useful as it allow assessments and analyses of the air quality impacts in a specific city.

In addition to that, an emission VRP is proposed by Figliozzi (2010) to incorporate minimization of both the economic costs and emissions. Time dependant VRP (TDVRP) is used in this research where the TDVRP uses links that have different constraints such as speed at peak hours. The author make use of a multi-objective function that includes distance travelled, route durations and emissions, together with a heuristic algorithm to solve several instances. Figliozzi (2010) proposed that it may be possible to reduce emissions with a minimal increase in routing costs.

Several literatures also look into the issue of emissions through analysis of travel times and CO_2 emissions done with a vehicle routing problem (VRP) model. A model consists of travel time, fuel and emissions is created by Jabali et al. (2013) and is solved via tabu search procedures. One very important point mentioned by them is the correlation between fuel consumption and CO_2 emissions, where reduction of emissions leads to cost reduction. Lower and upper bounds on the total emissions based on the VRP solutions are computed and quality of the numerical results is benchmarked against them. However, legal maximum working time of the driver is not considered in their research.

Zhu et al. (2014) proposed a fuel consumption minimization routing problem to solve for an environment friendly and cost effective route. They formulated an integer linear programming model which uses arc-elimination procedure to identify the optimal route. The routes are selected based on different safety factors. The control of vehicle emissions in their study is through constraining the vehicle routing distances. On the other hand, Kumar et al. (2015), proposed a modelling technique to deal with pollution-routing problems and evaluated the trade-off model by formulating it as a multi-objective multi-vehicle routing problem. Bi-objective model included objectives such as minimization of total cost and total emission of the routing problem. In addition, they proposed a hybrid self-learning Particle Swarm algorithm to obtain near optimal solution.

2.3 *Vehicle routing models and methods to reduce risk of hazardous material transportation*

One of the earliest green logistics research done in routing of hazardous material is by Zografos and Davis (1989). The proposed model includes routing risk, cost and property damages. The nature of the problem is a multi-objective problem, where a route which is the shortest, may not have the lowest risk. Decision making models are needed to cope with the different objectives that come with the routing problem. Later on, Zografos and Androutsopoulos (2004) proposed a heuristic algorithm for solving the hazardous material distribution problem. The insertion algorithm proposed by them builds the routes by inserting customers, one at a time, at each iteration. The insertion of unrouted points is also allowed so that reinsertion to a better position is possible.

Due to the risk involved, transportation of hazardous materials is a topic which attracted a number of researchers. Instead of solving the VRP for the shortest path, a trade-off between the risk and the distance should be considered. Tarantilis and Kiranoudis (2001) solved the VRP variant through the population exposure risk mitigation. The selection of routes is done in such a way that the route will not be close to aggregate population points to reduce the population exposure risk. Using this method, the number of people placed at risk in case of an accident is reduced. List Based Threshold Accepting

(LBTA) is a single stochastic search method, hence it is easy to tune. The algorithm iteratively searched the solution space for better solutions.

Leonelli et al. (2000) proposed a new methodology to select the best route for transporting hazardous material based on risk analysis. The analysis is based on node and arcs on a routing problem similar to VRP. Both the economic costs and risk related costs are taken into consideration in order to obtain the cheapest flow distribution. The optimization procedures based on linear risk sources and the costs are implemented on OptiPath, which is an optimization software.

Meng et al. (2005) proposed a novel vehicle routing and scheduling problem to transport hazardous materials using multi-objective concept. The space time network approach which is able to fully characterize feasible time varying path is employed to develop the solution. On top of that, the time varying multi-objective algorithm proposed is based on dynamic programming method. The algorithm is tested on a hypothetical shipment of gas in Singapore.

Sadjadi (2007) used the application of Efficient Frontier (EF) in solving the transportation of hazardous material problem. The method is able to provide sets of solutions which can be implemented by the decision-maker. The proposed method is modelled in convex quadratic optimization. However, this method may not be suitable for NP-hard problems.

In one of the recent research, Desai and Lim (2013) proposed a stochastic dynamic programming (SDP) approach to solve the routing problems. Conventional SDP requires long computational time and therefore three different techniques were proposed to expedite the process. The approach is applied on hazardous materials transportation problem. In terms of including risk, Faghih-Roohi et al. (2015) proposed a dynamic model for conditional value-at-risk (CVaR) of hazardous material transportation in the supply chain network. CVaR is a commonly used risk measure and is used as the main objective of their optimization problem. The effects of road conditions, type of hazards and other factors that are probable to accidents were also considered in their study. Recent study by Du et al. (2016) addresses the hazardous material transportation risk using chance-constrained programming modelling approach and credibility theory. Solution techniques proposed by them to determine near optimal solution combined both genetic algorithm and fuzzy simulation.

Overall the review reveals that few studies in the past modelled the two issues such as CO_2 emissions and hazardous material transportation problem separately and not considered them as a combined issue. Hence, we tried to address this gap in this study. The review also reveals that the multi-objective formulation is quite a feasible approach to deal with CO_2 reduction when there is more than one objective function. To the best of our knowledge, there is no safe vehicle routing analytical models and methods that reduces CO_2 and risk with a trade-off that will yield an efficient solution in the literature.

The major contributions of this study are to develop a multi-objective model for a safe vehicle routing and to propose a multi-objective differential evolution (MODE) algorithm that can handle structured and unstructured data to solve the safe capacitated vehicle routing problem.

3 Safe vehicle routing optimization considering CO_2 emission and hazardous risk

3.1 Multi-objective safe vehicle routing

VRP has been widely used for scheduling and planning routes in logistics. VRP framework aid the planning of least-cost delivery routes from one or more depot to a set of customers situated at different locations. In real world scenarios, the safe scheduling and planning of routes is subject to several constraints and customer requirements. The solution of a VRP consists of a set of routes with all the requirements and operational constraints satisfied while minimizing the transportation cost (Toth and Vigo 2002).

In order to incorporate environmental objectives into safe VRP, CO_2 emissions and hazardous materials risk are taken into consideration during development of route plans.

In general, VRP is represented by a graph $G = (V, A)$ where $V = \{1,2, \dots, N+1\}$ is the vertex set and A is a set of arcs. Vertex 1 denotes the depot and the rest of the vertex V, $\{2,3, \dots, N+1\}$ is referred to as customers. In a VRP model, the customers are serviced by a fleet of vehicles $\{1,2, \dots, K\}$. There is a specific demand, q for each customer. Safe CVRP, a classical VRP with additional constraint, which is used in this paper has a predefined capacity of Q for each vehicle. This is usually assumed to be the same for all vehicles in the set, unless otherwise stated.

Notation and Parameters

N	total number of customers
K	total number of vehicles
Q	maximum capacity of each vehicle
E_{ij}	CO_2 emission between customer i and j, d_{ij}=d_{ji}, $\forall i, j \in \{1,2, \dots, N+1\}$
r_{ij}	risk between customer i and j, d_{ij}=d_{ji}, $\forall i, j \in \{1,2, \dots, N+1\}$
q_i	demand at node i, ,$\forall i, \in \{1,2, \dots, N+1\}, q_1 = 0$

Decision variables

$$x_{ij}^k = \begin{cases} 1, if\ arc(i,j)\ belong\ to\ the\ route\ operated\ by\ vehicle\ k \\ 0, otherwise \end{cases}$$

$$x_{ij}^k \in \{0,1\}, i \neq j, j \in \{1,2, \dots, N+1\}$$

Formulations

The mathematical formulations of a multi-objective safe CVRP can be expressed as:

Minimize $\{Emission \quad , \quad Risk\ \}$

where $Emission$ and $Risk$ are defined in Section 3.2 and 3.3; and subject to

$$\sum_{k=1}^{K} \sum_{j=2}^{N+1} x_{ij}^k \leq K, for\ \ i = 1 \tag{1}$$

$$\sum_{j=2}^{N+1} x_{ij}^k = \sum_{j=2}^{N+1} x_{ji}^k, for\ i = 1\ and\ k \in \{1,2, \dots, K\} \tag{2}$$

$$\sum_{k=1}^{K} \sum_{j=1}^{N+1} x_{ij}^k = 1, for\ \ i \in \{2,3, \dots, N+1\} \tag{3}$$

$$\sum_{k=1}^{K} \sum_{i=1}^{N+1} x_{ij}^k = 1, for\ \ j \in \{2,3, \dots, N+1\} \tag{4}$$

$$\sum_{i=1}^{N+1} x_{ij}^k - \sum_{i=1}^{N+1} x_{ji}^k = 0, for\ j \in \{1,2,\dots,N+1\} and\ for\ k \in \{1,2,\dots,K\} \quad (5)$$

$$\sum_{i=1}^{N+1} q_i \left(\sum_{j=1}^{N+1} x_{ij}^k \right) \le Q, for\ k \in \{1,2,\dots,K\} \quad (6)$$

The objective functions of the multi-objective safe CVRP seek to minimize the total CO_2 emission (9) and the total risk of the vehicles transporting hazardous materials (19). Constraint (1) restricts the total number of vehicles at service to not exceed the maximum number of vehicles stated. In addition, equation (2) ensures that every route starts and ends at depot. Constraints (3) and (4) ensure that each customer node is only visited once. On top of that, constraint (5) guarantees that the same vehicle arrives and departs from each customer it serves and (6) ensures that the each vehicle do not load more that the vehicle capacity, Q. Equations (1) - (6) satisfy all the requirements of safe CVRP.

3.2 Emission function

An emission function in (7) is developed by the United Kingdom Transport Research Laboratory and reported in the MEET report (Hickman et al. 1999) and is also used in (Jabali et al. 2013; Demir et al. 2014). The function $\varepsilon(v)$ provides the rate of emission in g/km at travel speed v :

$$\varepsilon(v) = L + av + bv^2 + cv^3 + d\frac{1}{v} + e\frac{1}{v^2} + f\frac{1}{v^3} \quad (7)$$

where v is in km/h, and coefficients L and $a\ to\ f$ vary per vehicle type and size. Equation (7) is derived for heavy goods vehicles (HGVs), urban buses and coaches (Hickman et al. 1999).

Multiplication of rate of emissions (g/km) by the distances travelled (km) gives a total amount of CO_2 emission (in grams). Therefore, the CO_2 emission of a vehicle travels from customer i to customer j can be defined as:

$$E_{ij} = \varepsilon d_{ij} \quad (8)$$

where E_{ij} is in grams and d_{ij} is in kilometre. The total amount of emission for a route can be calculated as:

$$Emission = \sum_{k=1}^{K} \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} E_{ij} x_{ij}^k \quad (9)$$

As shown in equation (7), a minimum CO_2 emission is recorded when the vehicle is travelling at an optimal speed. However, in real life scenarios, this ideal condition may not be achieved particularly during the peak hours. Vehicles travelling on a path at that particular time will travel on a reduced speed. Since travel time is not a factor to be considered, the vehicle is assumed to travel in a constant average speed. Therefore, ε is constant in this paper.

Since the CO_2 emission in (8) is direct proportional to the distance between customer i and j, hence, minimizing the total amount of emission for a route can be simplified as minimizing the total distance travelled:

$$Distance = \sum_{k=1}^{K}\sum_{i=1}^{N+1}\sum_{j=1}^{N+1} d_{ij}x_{ij}^{k} \tag{10}$$

The optimal CO_2 emission can be calculated from the optimal distance using the following equation:

$$Emission = \varepsilon\sum_{k=1}^{K}\sum_{i=1}^{N+1}\sum_{j=1}^{N+1} d_{ij}x_{ij}^{k} = \varepsilon \times Distance \tag{11}$$

3.3 Hazardous Materials and Risk

Hazardous materials are defined as types of substances capable of causing harm or long term effects to the people, properties and the surrounding environment. However, various hazardous materials are used in today's industrialized societies and it is not possible to abstain from utilizing these materials (Kang et al. 2014). The United Nation (UN) sorted hazardous materials into nine different classes according to into its physical, chemical and nuclear properties (Erkut et al. 2007). The transportation of hazardous materials from one point to other is necessary and due to the risk involved, proper scheduling to reduce the transportation risk is vital. The vehicle routing problem with hazardous materials is about selecting a route that takes into consideration of economic and risk issues (Faghih-Roohi et al. 2015). In the process of planning the routes, the risk factors such as exposure and releases of dangerous material must be considered to achieve this objective. A route is considered to be safe if its risks are deemed to be acceptable (Alp 1995).

Risk estimation is a measure of the probability of harm to the exposed public and an assessment of the consequences of the undesirable incident. Hence, a risk of an event to the public must be addressed with both frequency and consequences components as (Alp 1995):

$$Event\ Risk = Frequency\ of\ occurence\ \times Estimated\ consequences \tag{12}$$

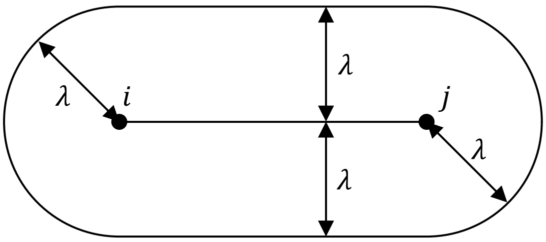


Figure 1: λ-neighbourhood of link (i,j)

The λ-neighbourhood concept shown in Figure 1 first developed by Batta and Chiu (1988). For travel on a link, the consequence is defined by assuming that the impact area is a danger circle with radius λ. Within this distance, the accident spots are subjected to the same consequence while the consequences outside this distance have been ignored. They described that a vehicle on point c of a link (i, j), poses a threat to point s if point c is within the radius λ of point s. The risk at point s due to hazardous materials transportation on link (i, j) is defined by:

$$r_{s,ij} = w_s \int_{c=0}^{d_{ij}} \delta(s,c)p_{ij}(c)\mathrm{d}c \tag{13}$$

where w_s is the population density at point s, $p_{ij}(c)$ is the probability of hazardous materials incident of vehicle at point c on link (i,j), and

$$\delta(s,c) = \begin{cases} 1, d_{sc} \leq \lambda \\ 0, \text{otherwise} \end{cases} \tag{14}$$

with d_{sc} is the Euclidean distance between point s and c. Denoted the integral term in (13) as:

$$F_{s,ij} = \int_{c=0}^{d_{ij}} \delta(s,c) p_{ij}(c) \mathrm{d}c \tag{15}$$

To simplify the equation, Batta and Chiu (1988) moved point i to the origin, and rotated the axes so that link (i,j) lies on x-axis. With point s having the Cartesian coordinates (x_s, y_s), x^+ and x^- are defined as the two intersections of link (i,j) with the circle of radius λ centered at point s:

$$x^+ = x_s + \sqrt{\lambda^2 - y_s{}^2} \text{ and } x^- = x_s - \sqrt{\lambda^2 - y_s{}^2}\text{, if } \lambda > |y_s| \tag{16}$$

They then identified the region within the λ-neighbourhood which lead to different expressions to compute $F_{s,ij}$:

$$F_{s,ij} = \begin{cases} & \text{when } s \text{ is} \\ \int_0^{x^+} p_{ij}(c)\mathrm{d}c, & \text{in region I} \\ \int_0^{d_{ij}} p_{ij}(c)\mathrm{d}c, & \text{in region II} \\ \int_{x^-}^{d_{ij}} p_{ij}(c)\mathrm{d}c, & \text{in region III} \\ \int_{x^-}^{x^+} p_{ij}(c)\mathrm{d}c, & \text{in region IV} \\ 0 & \text{outside of } \lambda\text{-neighbourhood} \end{cases} \tag{17}$$

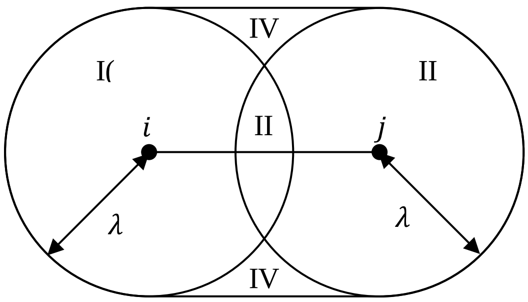


Figure 2: Regions inside the λ-neighbourhood of link (i,j)

As shown in (16), it can be seen that $x^- = x^+$ when point s is exactly located on the λ-boundary ($\lambda = |y_s|$), i.e. link (i,j) is tangent to the λ circle centred at point s. This intersect point is normally called the tangent point and can be seen in Figure 2. In this case, point c only poses a threat to point s if it is on the tangent point. When this occurs, it is shown that $F_{s,ij} = 0$ despite of the region point s is in. This can be proved using equation (17). If point s is in region I or III, the tangent point will be on point i and j relatively, therefore, $F_{s,ij} = 0$. If point s is in region IV and since $x^- = x^+$, the integral will be zero, hence $F_{s,ij} = 0$. Region II will not be considered as point s will never fall in region II in this case.

The total risk of a vehicle transporting hazardous materials and travels on link (i,j) is calculated by defining:

$$r_{ij} = \sum_{s=1}^{N+1} r_{s,ij} \tag{18}$$

And the total risk of vehicles transporting hazardous materials on a specific route is defined as:

$$Risk = \sum_{k=1}^{K} \sum_{i=1}^{N+1} \sum_{j=1}^{N+1} r_{ij} x_{ij}^{k} \tag{19}$$

4 Data Driven Multi-Objective Differential Evolution Algorithm

Data driven multi-objective problems take into consideration several criteria for decision making and optimal decisions are made between a certain trade-off based on real time data (Pérez et al. 2015).In literature, there are two ways to solve multi-objective optimization: by using weighted sum and to determine a set of Pareto optimal solutions. The weighted sum method is more straight-forward but the drawback lies on selecting the precise weight for the problem (Konak et al. 2006). In most real life engineering problems, a single solution does not exist and the trade-off between several objectives requires thorough analysis to make a certain compromise in order to obtain a better solution.

In this paper, the data driven multi-objective differential evolution (MODE) algorithm proposed involves a combination of Pareto ranking and crowding distance which is based on the improved differential evolution with local search (DELS) algorithm proposed by Teoh et al. (2015). The methods used are explained in the following sections.

4.1 Pareto Ranking

The concept of Pareto ranking is widely used in multi-objective algorithm to evaluate the fitness of a solution. The population is ranked according to a dominance rule, and each solution is assigned a fitness rank in the population (Goldberg 1989). Pareto ranking can be easily applied into the fitness evaluation process within an algorithm by replacing the fitness score with Pareto ranks (Ombuki et al. 2006), for which the lower ranks are always preferable.

Coello et al. (2002) discussed in their research about the idea of Pareto dominance and Pareto optimality which are part of Pareto ranking. The following definitions explain Pareto dominance, Pareto optimality and Pareto front.

Definition. Given a problem defined by a vector of objectives $f(s) = (f_1(s), \dots, f_m(s))$ subject to appropriate problem constraints, where s is a feasible solution. Then solution s_1 is said to dominate s_2 (denoted as $s_1 \prec s_2$) iff $\forall i \in (1, \dots, m)$: $f_i(s_1) \leq f_i(s_2)$ and $\exists i \in (1, \dots, m)$: $f_i(s_1) < f_i(s_2)$.

Definition. A solution s_2 is **Pareto Optimal** if there does not exist another solution s_1 that dominates s_2.

Definition. The **Pareto Optimal Set**, $\mathcal{P}$ is the set of all Pareto Optimal solutions defined by: $\mathcal{P} = \{s|s \text{ is Pareto Optimal}\}$.

Definition. The **Pareto Front** $\mathcal{PF}$ is defined by $\mathcal{PF} = \{s|s \in \mathcal{P}\}$.

The pseudo code of the Pareto ranking technique is shown in Algorithm 1. Pareto ranking, $\{\mathcal{F}_1, \mathcal{F}_2, \dots\}$ are called ***non-dominated fronts*** and $\mathcal{F}_1$ is the Pareto front of the generation.

Algorithm 1 Pareto Ranking

```
procedure PARETO_RANKING(U, f)
    i ← 1; S ← U
    repeat
        F[i] ← ∅
        for p ← 1, NP do
            if S[p] is non-dominated then
                F[i] ← F[i] ∪ S[p]
                PR[p] ← i
            end if
        end for
        S[S ∈ F[i]] ← ∅
        i ← i + 1
    until S = ∅
    return F and PR
end procedure
```

4.2 *Crowding Distance*

In the proposed data driven MODE, crowding distance is used as a tiebreaker in the selection phase. Crowding distance is an estimate of the density of the solutions surrounding a particular solution in a population (Deb et al. 2002). It is represented by the average distance of two points on either side of the particular solution along each objective function. Crowding distance method is chosen because it can be calculated without a user-defined parameter (Konak et al. 2006).

During selection, two solutions are selected for tournament. The solution with the lowest Pareto rank is the winner. However, if the solutions are in the same non-dominated front, the solution with a higher crowding distance is selected.

Algorithm 2 outlines the calculation procedure for the crowding distance for one non-dominated front, $\mathcal{F}$ of l solutions.

Algorithm 2 Crowding Distance

```
procedure CROWDING_DISTANCE(F, f)
    Cd ← 0
    for i ← 1,2 do
        [sf, I] ← sort(f[f ∈ F], i)
        Cd[I[1]] = Cd[I[l]] ← ∞
        for j ← 2, l − 1 do
            Cd[I[j]] ← Cd[I[j]] + (sf[j + 1] − sf[j − 1])/(max(f[i]) − min(f[i]))
        end for
    end for
    return Cd
end procedure
```

5 Computational Results

The proposed data driven MODE algorithm is coded and executed in MATLAB 7.11.0. In this section, the data driven MODE algorithm is tested with safe CVRP instances of Augerat et al. (1995) with the parameters and characteristics discussed in the following sections.

In the Augerat dataset, there are 74 instances and are categorised into three different sets, i.e. set A, B and P. The customer locations are randomized in both set A and P while clustered in set B. The capacity of the vehicle is constant in set A and B, whereas in set P, the capacity of the vehicle varies for each instance.

These instances are widely used and are publicly available at (Dorronsoro 2005). However to the best of our knowledge, these instances have not been used to minimize risk and emission.

5.1 Parameter Setup

The parameters used in data driven MODE are summarised in Table 1 below.

Table 1: Parameters used in data driven MODE

Parameters	Notation	Values	Unit	Reference
Radius of danger circle	λ	0.8	km	(DoT 1996)
Population density	w_s	1000	People	-
Release probability	p_{ij}	4.8×10^{-7}	per km	(Harwood et al. 1993)
Vehicle type	-	16 – 32 tonnes	-	-
Coefficients for vehicle type	$L, a - f$	{765, -7.04, 0, 0.006320, 8334, 0, 0}	-	(Hickman et al. 1999)
Travel speed	v	60	km/h	-

The impact radius, λ for Class 3 and Class 4 hazardous materials, which include flammable gasses, flammable or combustible liquids, flammable solids and spontaneously combustible materials have a potential impact area of $0.8km$ radius in all directions. This is chosen from Table 2 which is adopted from the 1996 Hazardous Materials Routing Guidelines (DoT 1996).

Table 2: Radius λ of impact area by hazardous materials class

Class	Hazardous Materials	Code	Radius λ
Class 1	Explosives	EXP	1.0 mi. (1.6 km) all directions
Class 2	Flammable Gas	FG	0.5 mi. (0.8 km) all directions
	Poison Gas	PG	5.0 mi. (8.0 km) all directions
Class 3	Flammable/Combustible Liquid	FCL	0.5 mi. (0.8 km) all directions
Class 4	Flammable Solid; Spontaneously Combustible, Dangerous when Wet	FS	0.5 mi. (0.8 km) all directions
Class 5	Oxidizer/Organic Peroxide	OXI	0.5 mi. (0.8 km) all directions
Class 6	Poisonous, not gas	POI	5.0 mi. (8.0 km) all directions
Class 8	Corrosive Material	COR	0.5 mi. (0.8 km) all directions

The estimation for the population density is a challenging task due to the fact that population in a location varies depends on the time of the day (Erkut and Verter 1998). The distribution in an area is usually not uniform. However, due to the lack of information, the variation in population is ignored. It is assumed that the population density at each customer node is 1000 people. Since the datasets are scaled to simulate an urban area, and all links are assumed to be multilane, the release probability, p_{ij} is assumed to be a constant of 4.8×10^{-7} per km (Harwood et al. 1993). The vehicle type selected for this research is the heavy goods vehicles with 16 – 32 tonnes of gross vehicle weight. The coefficients of CO_2 emission for different vehicle types are available at (Hickman et al. 1999). With consideration that the vehicle is transporting hazardous material and roads in an urban area are likely congested, it is assumed that the vehicle travels in an average speed of $60 km/h$.

The proposed MODE is based on the DELS algorithm, and therefore the parameters for DELS are taken from (Teoh et al. 2015) and presented in Table 3.

Table 3: Parameters and constants for MODE

Parameters	**Value**
Number of population, NP	$3N$
Crossover rate, Cr	0.4
Mutation scale factor, F	Randomized(0.5,1)

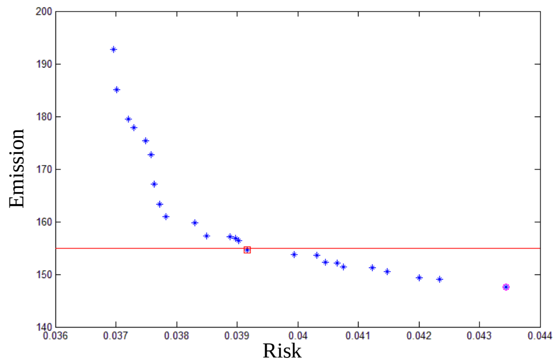
5.2 Analysis of Results

In this section, the results of the computational experiments of the proposed data driven MODE algorithm on CVRP instances are presented and analysed. The selection of the suitable optimal solution from the Pareto Front, $\mathcal{PF}$, is also explained in this section.

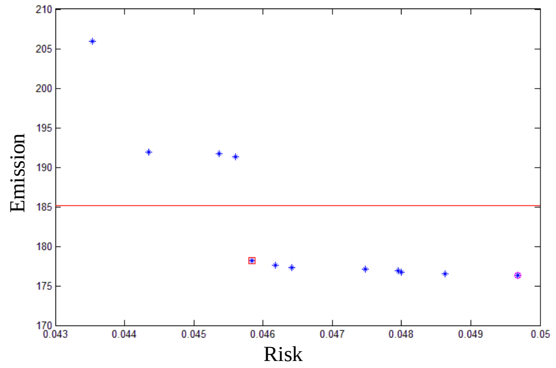
Theoretically, the Pareto Front contains the optimal solution set. A suitable solution based on the specific application has to be chosen. In this paper, a solution in the Pareto Front has to be chosen such that it has a lower risk and at the same time having a relatively low cost.

The pareto front generated by data driven MODE for three example instances are shown in Figure 3. The trade-off allowed in this paper is set to a 5% increase of emission cost as shown by the horizontal line in Figure 3. The circle marker is used for DELS solution and asterisk markers are the Pareto Front solutions. The optimal solution will be chosen from the set of solution in the Pareto Front which falls under 5% increment of emission cost as compared to the optimal solution found in DELS (Teoh et al. 2015). The solution with the lowest risk within that allowable range is chosen as the solution and is plotted using the square marker.

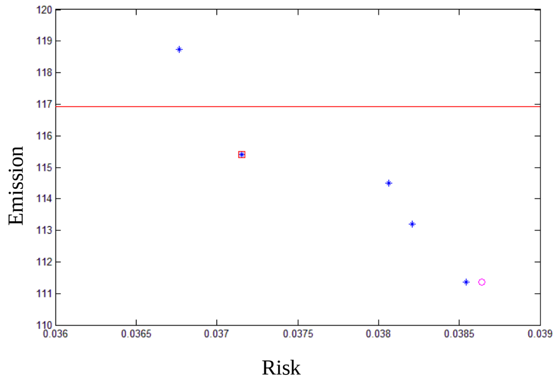
The percentage decrease in risk of this optimal solution is compared to the optimal solution found using DELS algorithm. All the results are tabulated in Table 4 to Table 6.



(a)



(b)



(c)

Figure 3: Pareto Front for (a) A-n36-k5 (b) B-n36-k5 (c) P-n22-k8

Table 4: Optimal Solution for Set A instances

Dataset	$\mathcal{PF}$	**DELS**			**Data driven MODE**			**Under 5% Cost Increase**	**Change in Risk**	**Ratio**
		Distance	**Emission**	**Risk**	**Distance**	**Emission**	**Risk**			
A-n32-k5	15	78.4	144.78	0.0366	82.2	151.79	0.0334	4.84%	-8.90%	1.84
A-n33-k5	12	66.1	122.06	0.0377	68.8	127.05	0.0341	4.09%	-9.55%	2.34
A-n33-k6	9	74.2	137.02	0.0437	76.9	142.01	0.0415	3.64%	-5.17%	1.42
A-n34-k5	12	77.8	143.67	0.0359	81.1	149.76	0.0341	4.24%	-4.93%	1.16
A-n36-k5	25	79.9	147.54	0.0434	83.7	154.56	0.0392	4.76%	-9.83%	2.07
A-n37-k5	22	66.9	123.54	0.0431	70.0	129.26	0.0393	4.63%	-8.71%	1.88
A-n37-k6	14	94.9	175.24	0.0482	99.4	183.55	0.0435	4.74%	-9.65%	2.03
A-n38-k5	10	73.0	134.80	0.0439	75.1	138.68	0.0393	2.88%	-10.61%	3.69
A-n39-k5	16	82.2	151.79	0.0511	85.4	157.70	0.0473	3.89%	-7.41%	1.90
A-n39-k6	19	83.1	153.45	0.0469	86.5	159.73	0.0423	4.09%	-9.83%	2.40
A-n44-k6	22	93.7	173.03	0.0545	98.0	180.97	0.0498	4.59%	-8.63%	1.88
A-n45-k6	10	94.4	174.32	0.0566	97.9	180.78	0.0506	3.71%	-10.51%	2.84
A-n45-k7	7	114.6	211.62	0.0684	117.4	216.79	0.0617	2.44%	-9.76%	3.99
A-n46-k7	14	91.4	168.78	0.0621	94.6	174.69	0.0529	3.50%	-14.84%	4.24
A-n48-k7	17	107.3	198.14	0.0599	111.9	206.64	0.0557	4.29%	-6.90%	1.61
A-n53-k7	19	101.0	186.51	0.0669	104.0	192.05	0.0603	2.97%	-9.75%	3.28
A-n54-k7	13	116.7	215.50	0.0807	122.4	226.03	0.0754	4.89%	-6.54%	1.34
A-n55-k9	18	107.3	198.14	0.0809	112.2	207.19	0.0771	4.57%	-4.68%	1.03
A-n60-k9	11	135.4	250.03	0.0985	141.9	262.04	0.0947	4.80%	-3.90%	0.81
A-n61-k9	10	103.5	191.13	0.1001	108.5	200.36	0.0931	4.83%	-6.95%	1.44
A-n62-k8	14	128.8	237.84	0.0980	135.1	249.48	0.0917	4.89%	-6.42%	1.31
A-n63-k9	4	162.4	299.89	0.1049	165.0	304.69	0.0933	1.60%	-11.12%	6.95
A-n63-k10	13	131.6	243.02	0.1061	137.4	253.73	0.0946	4.41%	-10.90%	2.47
A-n64-k9	11	141.6	261.48	0.1158	148.6	274.41	0.1000	4.94%	-13.65%	2.76
A-n65-k9	8	118.1	218.09	0.0789	122.5	226.21	0.0739	3.72%	-6.32%	1.70
A-n69-k9	11	116.5	215.13	0.0973	122.0	225.29	0.0891	4.72%	-8.39%	1.78
A-n80-k10	4	177.9	328.51	0.1244	185.5	342.55	0.1163	4.27%	-6.49%	1.52
Average		104.4	192.78	0.0698	108.7	200.67	0.0639	4.11%	-8.53%	-

Table 5: Optimal Solution for Set B instances

Dataset	$\mathcal{PF}$	DELS			Data driven MODE			Under 5% Cost Increase	Change in Risk	Ratio
		Distance	Emission	Risk	Distance	Emission	Risk			
B-n31-k5	5	67.2	124.09	0.0603	67.9	125.39	0.0570	1.05%	-5.42%	5.18
B-n34-k5	3	78.8	145.51	0.0510	82.6	152.53	0.0492	4.82%	-3.67%	0.76
B-n35-k5	12	95.5	176.35	0.0497	96.5	178.20	0.0458	1.05%	-7.73%	7.37
B-n38-k6	13	80.5	148.65	0.0578	84.2	155.49	0.0555	4.60%	-3.98%	0.86
B-n39-k5	14	54.9	101.38	0.0712	56.7	104.70	0.0663	3.27%	-6.94%	2.12
B-n41-k6	20	82.9	153.08	0.0711	84.6	156.22	0.0666	2.05%	-6.41%	3.12
B-n43-k6	17	74.2	137.02	0.0816	77.8	143.67	0.0658	4.85%	-19.41%	4.00
B-n44-k7	17	90.9	167.86	0.1060	95.3	175.98	0.0943	4.84%	-11.04%	2.28
B-n45-k5	13	75.1	138.68	0.0698	78.5	144.96	0.0644	4.53%	-7.76%	1.71
B-n45-k6	14	67.8	125.20	0.0937	71.0	131.11	0.0869	4.72%	-7.26%	1.54
B-n50-k7	19	74.1	136.83	0.0840	75.0	138.50	0.0778	1.22%	-7.37%	6.04
B-n50-k8	18	131.3	242.46	0.1431	135.8	250.77	0.1172	3.43%	-18.11%	5.29
B-n51-k7	11	103.3	190.76	0.0864	106.7	197.03	0.0801	3.29%	-7.28%	2.21
B-n52-k7	16	74.7	137.94	0.0877	78.1	144.22	0.0825	4.55%	-5.92%	1.30
B-n56-k7	4	70.7	130.56	0.1105	71.5	132.03	0.1063	1.13%	-3.82%	3.39
B-n57-k7	2	116.6	215.32	0.1188	117.1	216.24	0.1116	0.43%	-6.02%	14.09
B-n57-k9	17	159.9	295.27	0.1205	167.5	309.31	0.1069	4.75%	-11.28%	2.37
B-n63-k10	13	150.4	277.73	0.1161	157.8	291.40	0.1095	4.92%	-5.66%	1.15
B-n64-k9	9	86.1	158.99	0.1311	89.7	165.64	0.1255	4.18%	-4.32%	1.03
B-n66-k9	11	132.2	244.12	0.1570	135.9	250.96	0.1384	2.80%	-11.84%	4.22
B-n67-k10	15	103.2	190.57	0.1318	108.2	199.80	0.1239	4.84%	-5.97%	1.23
B-n68-k9	19	128.1	236.55	0.1502	133.5	246.52	0.1366	4.21%	-9.08%	2.15
B-n78-k10	10	123.0	227.13	0.1667	128.9	238.03	0.1534	4.80%	-7.95%	1.66
Average		96.6	178.35	0.1007	100.0	184.73	0.0922	3.49%	-8.01%	-

Table 6: Optimal Solution for Set P instances

Dataset	$\mathcal{PF}$	DELS			Data driven MODE			Under 5% Cost Increase	Change in Risk	Ratio
		Distance	**Emission**	**Risk**	**Distance**	**Emission**	**Risk**			
P-n16-k8	1	45.0	83.10	0.0347	45.0	83.10	0.0347	0.00%	0.00%	-
P-n19-k2	1	21.2	39.15	0.0198	21.2	39.15	0.0198	0.00%	0.00%	-
P-n20-k2	3	21.6	39.89	0.0212	21.8	40.26	0.0204	0.93%	-3.87%	4.18
P-n21-k2	2	21.1	38.96	0.0220	21.6	39.89	0.0219	2.39%	-0.41%	0.17
P-n22-k2	1	21.6	39.89	0.0251	21.6	39.89	0.0251	0.00%	0.00%	-
P-n22-k8	5	60.3	111.35	0.0386	62.5	115.41	0.0372	3.65%	-3.86%	1.06
P-n23-k8	2	53.3	98.42	0.0671	53.6	98.98	0.0660	0.57%	-1.58%	2.78
P-n40-k5	2	45.8	84.58	0.0465	45.9	84.76	0.0463	0.21%	-0.52%	2.42
P-n45-k5	1	51.0	94.18	0.0525	51.0	94.18	0.0525	0.00%	0.00%	-
P-n50-k7	4	55.4	102.30	0.0753	55.7	102.86	0.0721	0.55%	-4.21%	7.69
P-n50-k8	1	64.1	118.37	0.0863	64.1	118.37	0.0863	0.00%	0.00%	-
P-n50-k10	7	69.6	128.52	0.0967	71.1	131.29	0.0945	2.16%	-2.29%	1.06
P-n51-k10	2	74.2	137.02	0.0993	74.8	138.13	0.0977	0.81%	-1.59%	1.96
P-n55-k7	6	56.8	104.89	0.0859	58.2	107.47	0.0785	2.46%	-8.66%	3.52
P-n55-k8	5	58.9	108.77	0.0840	59.9	110.61	0.0829	1.69%	-1.31%	0.77
P-n55-k10	8	69.4	128.16	0.1051	70.7	130.56	0.1011	1.87%	-3.79%	2.02
P-n55-k15	1	98.9	182.63	0.1529	98.9	182.63	0.1529	0.00%	0.00%	-
P-n60-k10	7	74.4	137.39	0.1117	77.4	142.93	0.1087	4.03%	-2.67%	0.66
P-n60-k15	3	96.8	178.75	0.1485	98.1	181.15	0.1458	1.34%	-1.78%	1.32
P-n65-k10	3	79.2	146.25	0.1175	80.4	148.47	0.1140	1.52%	-3.02%	1.99
P-n70-k10	3	82.7	152.72	0.1382	85.6	158.07	0.1326	3.50%	-4.09%	1.17
P-n76-k4	1	59.3	109.50	0.0961	59.3	109.50	0.0961	0.00%	0.00%	-
P-n76-k5	1	62.9	116.15	0.1059	62.7	115.78	0.1020	-0.32%	-3.67%	N/A
P-n101-k4	1	68.5	126.49	0.1400	68.4	126.31	0.1380	-0.14%	-1.48%	N/A
Average		58.8	108.64	0.0821	59.6	109.99	0.0803	1.13%	-2.03%	-

The first column in each table signifies the set instance's name and the second column represents the number of solution in Pareto Front. The next 2 columns, each with 3 sub-columns, represent the results (distance, emission and risk) of optimal solutions from DELS algorithm and the chosen solution from the Pareto Front of the proposed data driven MODE algorithm. The last three columns respectively show the percentage of increase in cost, percentage of decrease in risk and the ratio of these two percentages. The positive values signify an increase whereas negative values indicate a decrease. These values are computed using Equations (20) and (21).

$$\%Increase \text{ or } \%Decrease = \frac{(Cost_{MODE} - Cost_{DELS})}{Cost_{DELS}} \times 100\% \qquad (20)$$

$$\text{Ratio} = \left|\frac{\%Decrease}{\%Increase}\right| \qquad (21)$$

From the percentage decrease in risk shown in the tables, the proposed data driven MODE algorithm provides a good range of Pareto solution for the multi-objective problem. The chosen optimal solution from the Pareto Front is more efficient compared to the optimal solutions found in (Teoh et al. 2015).

The optimal solution will be chosen from the Pareto Front if it has lesser than 5% increment of emission cost from the DELS solutions. A comparison is done for the percentage of risk increment between the DELS solutions and the solutions from data driven MODE.

For set A instances, it can be seen from Table 4 that the chosen solutions have an average of 8.53% decrease in risk while the cost increased by 4.11% in average. The maximum decrease in risk is 14.84% and the minimum percentage of decrement is 3.90%. The particular instance with a decrement of 14.84% in risk has a low increment of cost by 3.50%. This result shows that it is possible to achieve a huge risk reduction with minimal increase in cost. The highest ratio of decrement of risk to increment of cost for set A is at 6.95 while the minimum is at 0.81.

The chosen solutions of set B instances managed to achieve an average of 8.01% decrease in risk and 3.49% increase in cost. From Table 5, B-n43-k6 has the highest percentage of decrement in risk at 19.41%. With a little increase in cost (4.85%), a relatively high ratio of decrement to increment is achieved. The highest ratio for the B instances is at 14.09. This shows that the proposed algorithm performed well in clustered datasets.

As shown in Table 6, the proposed algorithm successfully provided a range of solutions for the 15 instances, whereas the remaining 9 instances have only one solution in the Pareto Front. No trade-offs are required for the 9 instances which have only a single solution as the lowest cost route possesses the lowest risk. The set P instances achieved an average of 2.03% decrease in risk and 1.13% increase in cost. The highest decrement in risk is at 8.66% and this particular instance has 2.46% increment in cost.

Benchmarking for the results found using data driven MODE is done with the Non-dominated Sorting-based Genetic Algorithm II (NSGA-II). NSGA-II is a well-known multi-objective optimization algorithm. The comparison is done based on the NSGA-II source code which is publicly available online (Deb 2008). The existing source code is modified to adopt the multi-objective optimization problems used in this paper. The results are tabulated in Table 7 to Table 9. The data MODE algorithm is found to be competitive to NSGA-II and for certain data sets, the data driven MODE results are found to be better than the NSGA-II. For set-P instances, the proposed algorithm obtained solutions with lower risks as compared to NSGA-II.

Table 7: Benchmark for set A instances

Dataset	Data driven MODE			NSGA-II		
	Under 5% Cost Increase	Change in Risk	Ratio	Under 5% Cost Increase	Change in Risk	Ratio
A-n32-k5	4.84%	-8.90%	1.84	2.80%	-7.97%	2.84
A-n33-k5	4.09%	-9.55%	2.34	2.12%	-9.18%	4.33
A-n33-k6	3.64%	-5.17%	1.42	3.64%	-5.17%	1.42
A-n34-k5	4.24%	-4.93%	1.16	4.24%	-4.93%	1.16
A-n36-k5	4.76%	-9.83%	2.07	4.26%	-8.06%	1.89
A-n37-k5	4.63%	-8.71%	1.88	3.44%	-7.25%	2.11
A-n37-k6	4.74%	-9.65%	2.03	4.74%	-9.65%	2.03

A-n38-k5	2.88%	-10.61%	3.69	4.38%	-10.61%	2.42
A-n39-k5	3.89%	-7.41%	1.90	3.89%	-7.41%	1.90
A-n39-k6	4.09%	-9.83%	2.40	4.09%	-9.83%	2.40
A-n44-k6	4.59%	-8.63%	1.88	0.21%	1.32%	6.18
A-n45-k6	3.71%	-10.51%	2.84	4.98%	-10.27%	2.06
A-n45-k7	2.44%	-9.76%	3.99	4.80%	-10.81%	2.25
A-n46-k7	3.50%	-14.84%	4.24	3.94%	-15.39%	3.91
A-n48-k7	4.29%	-6.90%	1.61	4.29%	-6.90%	1.61
A-n53-k7	2.97%	-9.75%	3.28	3.96%	-10.19%	2.57
A-n54-k7	4.89%	-6.54%	1.34	4.71%	-8.45%	1.79
A-n55-k9	4.57%	-4.68%	1.03	4.85%	-3.74%	0.77
A-n60-k9	4.80%	-3.90%	0.81	4.21%	-5.51%	1.31
A-n61-k9	4.83%	-6.95%	1.44	4.92%	-8.58%	1.74
A-n62-k8	4.89%	-6.42%	1.31	4.97%	-8.28%	1.67
A-n63-k9	1.60%	-11.12%	6.95	2.83%	-11.58%	4.09
A-n63-k10	4.41%	-10.90%	2.47	4.86%	-11.35%	2.34
A-n64-k9	4.94%	-13.65%	2.76	4.87%	-15.93%	3.27
A-n65-k9	3.72%	-6.32%	1.70	0.76%	-5.04%	6.63
A-n69-k9	4.72%	-8.39%	1.78	4.89%	-8.74%	1.78
A-n80-k10	4.27%	-6.49%	1.52	4.10%	-8.84%	2.15
Average	4.11%	-8.53%	2.28	3.92%	-8.46%	2.54
Min	1.60%	-14.84%	0.81	0.21%	-15.93%	0.77
Max	4.94%	-3.90%	6.95	4.98%	1.32%	6.63

Table 8: Benchmark for set B instances

Dataset	Data driven MODE			NSGA-II		
	Under 5% Cost Increase	**Decrease in Risk**	**Ratio**	**Under 5% Cost Increase**	**Decrease in Risk**	**Ratio**
B-n31-k5	1.05%	-5.42%	5.18	0.60%	-3.90%	6.54
B-n34-k5	4.82%	-3.67%	0.76	0.00%	0.00%	-
B-n35-k5	1.05%	-7.73%	7.37	1.05%	-7.73%	7.37
B-n38-k6	4.60%	-3.98%	0.86	4.60%	-3.98%	0.86
B-n39-k5	3.27%	-6.94%	2.12	3.27%	-6.94%	2.12
B-n41-k6	2.05%	-6.41%	3.12	1.81%	-5.06%	2.80
B-n43-k6	4.85%	-19.41%	4.00	2.83%	-18.47%	6.52
B-n44-k7	4.84%	-11.04%	2.28	4.73%	-10.91%	2.31
B-n45-k5	4.53%	-7.76%	1.71	3.99%	-8.73%	2.19
B-n45-k6	4.72%	-7.26%	1.54	3.10%	-6.09%	1.97
B-n50-k7	1.22%	-7.37%	6.04	1.49%	-6.97%	4.67
B-n50-k8	3.43%	-18.11%	5.29	4.95%	-18.18%	3.67
B-n51-k7	3.29%	-7.28%	2.21	0.00%	0.00%	-
B-n52-k7	4.55%	-5.92%	1.30	4.55%	-5.92%	1.30
B-n56-k7	1.13%	-3.82%	3.39	1.13%	-1.21%	1.08
B-n57-k7	0.43%	-6.02%	14.09	4.37%	-9.05%	2.07

B-n57-k9	4.75%	-11.28%	2.37	4.57%	-12.23%	2.68
B-n63-k10	4.92%	-5.66%	1.15	4.32%	-5.87%	1.36
B-n64-k9	4.18%	-4.32%	1.03	4.99%	-5.02%	1.00
B-n66-k9	2.80%	-11.84%	4.22	4.47%	-14.40%	3.23
B-n67-k10	4.84%	-5.97%	1.23	0.00%	0.00%	-
B-n68-k9	4.21%	-9.08%	2.15	4.21%	-9.17%	2.18
B-n78-k10	4.80%	-7.95%	1.66	0.00%	0.00%	-
Average	3.49%	-8.01%	3.26	2.96%	-7.26%	2.94
Min	0.43%	-19.41%	0.76	0.00%	-18.47%	0.86
Max	4.92%	-3.67%	14.09	4.99%	0.00%	7.37

Table 9: Benchmark for set P instances

Dataset	Data driven MODE			NSGA-II		
	Under 5% Cost Increase	Decrease in Risk	Ratio	Under 5% Cost Increase	Decrease in Risk	Ratio
P-n16-k8	0.00%	0.00%	-	0.00%	0.00%	-
P-n19-k2	0.00%	0.00%	-	0.00%	0.00%	-
P-n20-k2	0.93%	-3.87%	4.18	0.00%	0.00%	-
P-n21-k2	2.39%	-0.41%	0.17	2.39%	-0.41%	0.17
P-n22-k2	0.00%	0.00%	-	0.00%	0.00%	-
P-n22-k8	3.65%	-3.86%	1.06	3.65%	-3.86%	1.06
P-n23-k8	0.57%	-1.58%	2.78	0.00%	0.00%	-
P-n40-k5	0.21%	-0.52%	2.42	0.00%	0.00%	-
P-n45-k5	0.00%	0.00%	-	0.00%	0.00%	-
P-n50-k7	0.55%	-4.21%	7.69	0.55%	-4.21%	7.69
P-n50-k8	0.00%	0.00%	-	0.00%	0.00%	-
P-n50-k10	2.16%	-2.29%	1.06	0.00%	0.00%	-
P-n51-k10	0.81%	-1.59%	1.96	3.23%	-0.78%	0.24
P-n55-k7	2.46%	-8.66%	3.52	2.11%	-5.87%	2.78
P-n55-k8	1.69%	-1.31%	0.77	0.00%	0.00%	-
P-n55-k10	1.87%	-3.79%	2.02	1.72%	-3.88%	2.25
P-n55-k15	0.00%	0.00%	-	0.00%	0.00%	-
P-n60-k10	4.03%	-2.67%	0.66	2.69%	-0.95%	0.35
P-n60-k15	1.34%	-1.78%	1.32	0.00%	0.00%	-
P-n65-k10	1.52%	-3.02%	1.99	1.89%	-2.16%	1.14
P-n70-k10	3.50%	-4.09%	1.17	3.63%	-2.88%	0.79
P-n76-k4	0.00%	0.00%	-	0.00%	0.00%	-
P-n76-k5	-0.32%	-3.67%	-11.53	0.00%	0.00%	-
P-n101-k4	-0.14%	-1.48%	-10.39	0.00%	0.00%	-
Average	1.13%	-2.03%	0.64	0.99%	-1.14%	1.83
Min	-0.32%	-8.66%	-11.53	0.00%	-5.87%	0.17
Max	4.03%	0.00%	7.69	3.65%	0.00%	7.69

6 Conclusion

The issues of safe and intelligent transportation system has been an on-going challenge to logistic companies, local governments, business owners, consumers and the population who will be directly or indirectly affected by the risks of the shipments. With the current developing industries, frequent logistics are required and to make matters worse, some of these shipments involve hazardous materials. Accident probabilities for a path may be low, but the undesirable effects of an accident will impose a great risk to the populations nearby. Thorough and proper planning of routes have been given priorities in order to reduce the risk to the nearby population and to achieve safety.

In this paper, a data driven multi-objective DE algorithm is introduced to optimize the two objectives defined for the safe CVRP problem. The classical CVRP problem is expanded to consider safer objectives. The optimization is done to reduce the hazardous material risk, CO_2 emission and at the same time to minimize the cost function. The data driven MODE algorithm incorporates DE algorithm with Pareto ranking and crowding distance techniques. The decision-maker then considers the trade-off to choose a solution from the set of optimal solutions in the Pareto Front. Computational results found proved the viability of the data driven MODE algorithm to solve the multi-objective problem with a certain trade-off to achieve an efficient and feasible route.

In this study we didn't consider travel time and assumed vehicle to travel in a constant average speed. In addition, we didn't consider traffic congestion issues. There is a potential avenue for future researchers to include real time issues using social media and other textual data. Moreover, variable speed profile and travel time information can be easily included in our data-driven multi-objective optimization problem. In terms of solution slighter modification of our MODE algorithm will suffice to include the total travel time. The quick convergence and adaptation of the DE approach will be an added advantage for real-time dynamic problems. A more comprehensive investigation of a wider range of real time challenges will make MODE as a rigorous tool to solve several variants of VRP.

Acknowledgements

We would like to thank the special issue editors for their support and encouragement. We are highly indebted to the reviewers for their constructive comments that substantially improved the quality and readability of our paper.

7 References

Alp, E. (1995). Risk-based transportation-planning practice-overall methodology and a case example. *INFOR, 33*(1), 4-19.

Augerat, P., Belenguer, J., Benavent, E., Corberán, A., Naddef, D., & Rinaldi, G. (1995). *Computational results with a branch and cut code for the capacitated vehicle routing problem*: IMAG.

Batta, R., & Chiu, S. S. (1988). Optimal obnoxious paths on a network: transportation of hazardous materials. *Operations Research, 36*(1), 84-92.

Brandenburg, M., & Rebs, T. (2015). Sustainable supply chain management: a modeling perspective. *Annals of Operations Research, 229*(1), 213-252.

Coello, C. A. C., Van Veldhuizen, D. A., & Lamont, G. B. (2002). *Evolutionary algorithms for solving multi-objective problems*. New York: Springer.

Davis, N., Lents, J., Osses, M., Nikkila, N., & Barth, M. (2005). Development and Application of an International Vehicle Emissions Model. *Transportation Research Record: Journal of the Transportation Research Board 1939*(1), 155-165.

Deb, K. (2008). Kanpur Genetic Algorithms Laboratory. http://www.iitk.ac.in/kangal/index.shtml. Accessed March 2015.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182-197.

Demir, E., Bektaş, T., & Laporte, G. (2014). The bi-objective pollution-routing problem. *European Journal of Operational Research, 232*(3), 464-478.

Desai, S., & Lim, G. J. (2013). Solution time reduction techniques of a stochastic dynamic programming approach for hazardous material route selection problem. *Computers & Industrial Engineering, 65*(4), 634-645.

Dorronsoro, B. (2005). The VRP Web. http://neo.lcc.uma.es/vrp/.

DoT, U. S. (1996). Highway Routing of Hazardous Materials Guidelines for Applying Criteria. Washington DC: National Highway Institute.

Du, J., Yu, L., & Li, X. (2016). Fuzzy multi-objective chance-constrained programming model for hazardous materials transportation. *International Journal of General Systems, 45*(3), 286-310.

EPA, U. S. (2006). Greenhouse Gas Emissions from the U.S Transportation Sector.

EPA, U. S. (2014). Inventory of U.S. Greenhouse Gas Emissions and Sinks : 1990-2012.

Erkut, E., Tjandra, S. A., & Verter, V. (2007). Hazardous Material Transportation. In *Handbooks in operations research and management science* (pp. 539-621).

Erkut, E., & Verter, V. (1998). Modelling of transport risk for hazardous materials. *Operations Research, 46*(5), 625-642.

Faghih-Roohi, S., Ong, Y.-S., Asian, S., & Zhang, A. N. (2015). Dynamic conditional value-at-risk model for routing and scheduling of hazardous material transportation networks. *Annals of Operations Research*, 1-20.

Figliozzi, M. (2010). Vehicle routing problem for emissions minimization. *Transportation Research Record: Journal of the Transportation Research Board, 2197*(1), 1-7.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning* (Vol. 412). Reading Menlo Park: Addison-wesley.

Harwood, D. W., Viner, J. G., & Russell, E. R. (1993). Procedure for developing truck accident and release rates for hazmat routing. *Journal of Transportation Engineering, 119*(2), 189-199.

Hazen, B. T., Skipper, J. B., Boone, C. A., & Hill, R. R. (2016). Back in business: operations research in support of big data analytics for operations and supply chain management. *Annals of Operations Research*, 1-11.

Hickman, J., Hassel, D., Joumard, R., Samaras, Z., & Sorenson, S. (1999). Methodology for calculating transport emissions and energy consumption. Transport Research Laboratory.

Jabali, O., Woesel, T. V., & Kok, A. G. d. (2013). Analysis of Travel Times and CO2 Emissions in Time-Dependent Vehicle Routing. *Production and Operations Management, 21*(6), 1060-1074.

Kang, Y., Batta, R., & Kwon, C. (2014). Value-at-risk model for hazardous material transportation. *Annals of Operations Research, 222*(1), 361-387.

Konak, A., Coit, D. W., & Smith, A. E. (2006). Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety, 91*(9), 992-1007.

Kumar, R. S., Kondapaneni, K., Dixit, V., Goswami, A., Thakur, L., & Tiwari, M. (2015). Multi-objective modeling of production and pollution routing problem with time window: A self-learning particle swarm optimization approach. *Computers & Industrial Engineering*.

Leonelli, P., Bonvicini, S., & Spadoni, G. (2000). Hazardous materials transportation: a risk-analysis-based routing methodology. *Journal of Hazardous Materials, 71*(1), 283-300.

Meng, Q., Lee, D.-H., & Cheu, R. L. (2005). Multiobjective vehicle routing and scheduling problem with time window constraints in hazardous material transportation. *Journal of transportation engineering, 131*(9), 699-707.

OECD (2015). Big Data and Transport Understanding and assessing options

Ombuki, B., Ross, B. J., & Hanshar, F. (2006). Multi-objective genetic algorithms for vehicle routing problem with time windows. *Applied Intelligence, 24*(1), 17-30.

Pérez, J. C., Carrillo, M. H., & Montoya-Torres, J. R. (2015). Multi-criteria approaches for urban passenger transport systems: A literature review. *Annals of Operations Research, 226*(1), 69-87.

Rodrigue, J.-P. (2013). *The geography of transport systems*: Routledge.

Rodrigue, J.-P., Slack, B., & Comtois, C. (2001). Green Logistics (The Paradoxes of). In *The Handbook of Logistics and Supply-Chain Management 2*.

Sadjadi, S. J. (2007). An application of efficient frontier in transportation of hazardous materials. *Computers & Industrial Engineering, 53*(2), 357-360.

Savelsbergh, M., & Van Woensel, T. (2016). 50th Anniversary Invited Article—City Logistics: Challenges and Opportunities. *Transportation Science, 50*(2), 579-590.

Sinuany-Stern, Z., & Sherman, H. D. (2014). Operations research in the public sector and nonprofit organizations. *Annals of Operations Research, 221*(1), 1-8.

Tarantilis, C. D., & Kiranoudis, C. T. (2001). Using the vehicle routing problem for the transportation of hazardous materials. *Operational Research, 1*(1), 67-78.

Teoh, B. E., Ponnambalam, S., & Kanagaraj, G. (2015). Differential evolution algorithm with local search for capacitated vehicle routing problem. *International Journal of Bio-Inspired Computation, 7*(5), 321-342.

Toth, P., & Vigo, D. (2002). An overview of vehicle routing problems. *The vehicle routing problem, 9*, 1-26.

Verter, V. (2011). Hazardous Materials Transportation. In *Wiley Encyclopedia of Operations Research and Management Science*: John Wiley & Sons

Zhu, X., Garcia-Diaz, A., Jin, M., & Zhang, Y. (2014). Vehicle fuel consumption minimization in routing over-dimensioned and overweight trucks in capacitated transportation networks. *Journal of Cleaner Production, 85*, 331-336.

Zografos, K. G., & Androutsopoulos, K. N. (2004). A heuristic algorithm for solving hazardous materials distribution problems. *European Journal of Operational Research, 152*(2), 507-519.

Zografos, K. G., & Davis, C. F. (1989). Multi Programming Approach for routing hazardous materials. *Journal of Transportation Engineering, 115*(6), 661-673.